

Moderne en krachtige technieken voor verzekeraars: fraudedetectie

29 oktober 2003

Prof. dr. H.M.P.Kersten

Principal consultant LogicaCMG en

Hoogleraar Bedrijfswiskunde Vrije Universiteit Amsterdam

De afgelopen jaren kenmerken zich door:

1. Exponentiële ontwikkeling van rekenkracht van computers
2. Exponentiële ontwikkeling van geheugencapaciteit van computers
3. Nieuwe technieken die op bovenstaande bouwen

Wie beduvelt de boel, wie loopt langs de rand,



There are 11 human faces in the picture. Can you find them all ?

Normal people find 4 or 5 of them.

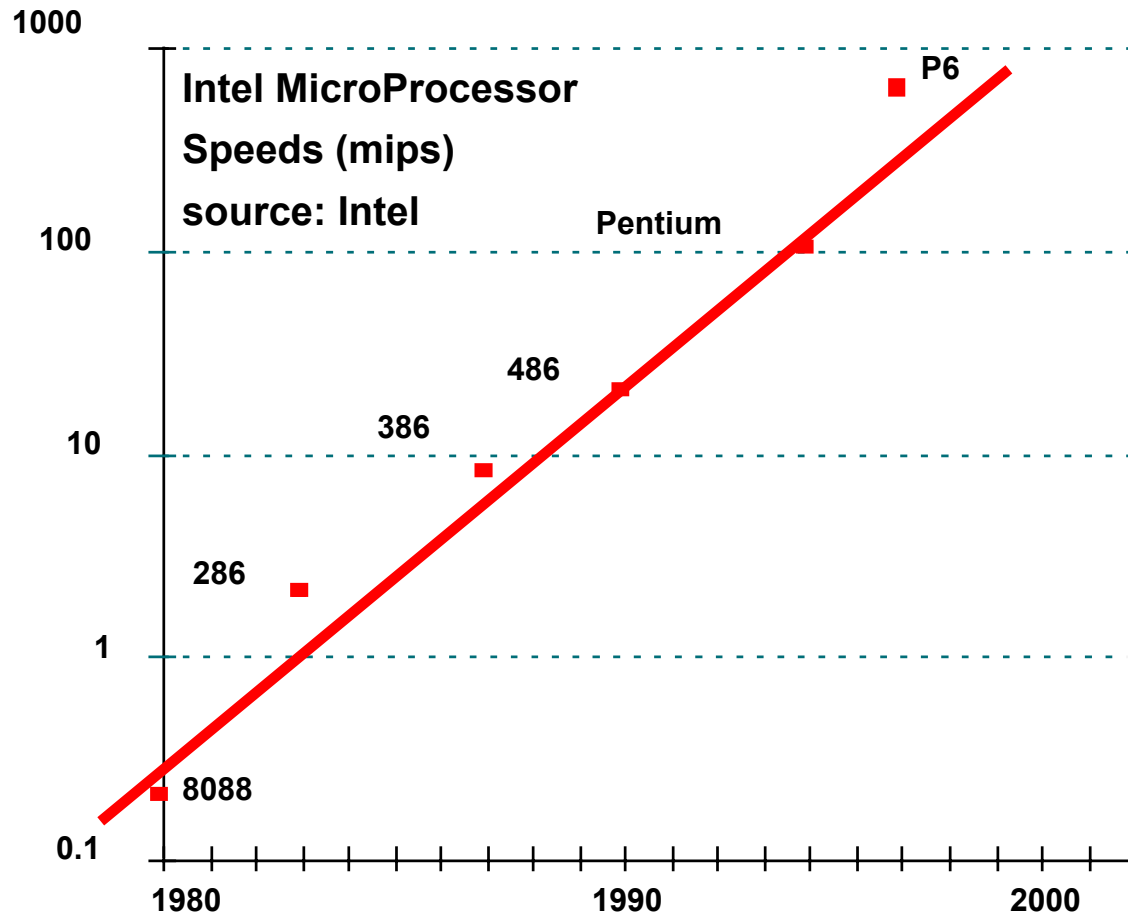
If you find 8 of them, you have a extraordinary sense of observation.



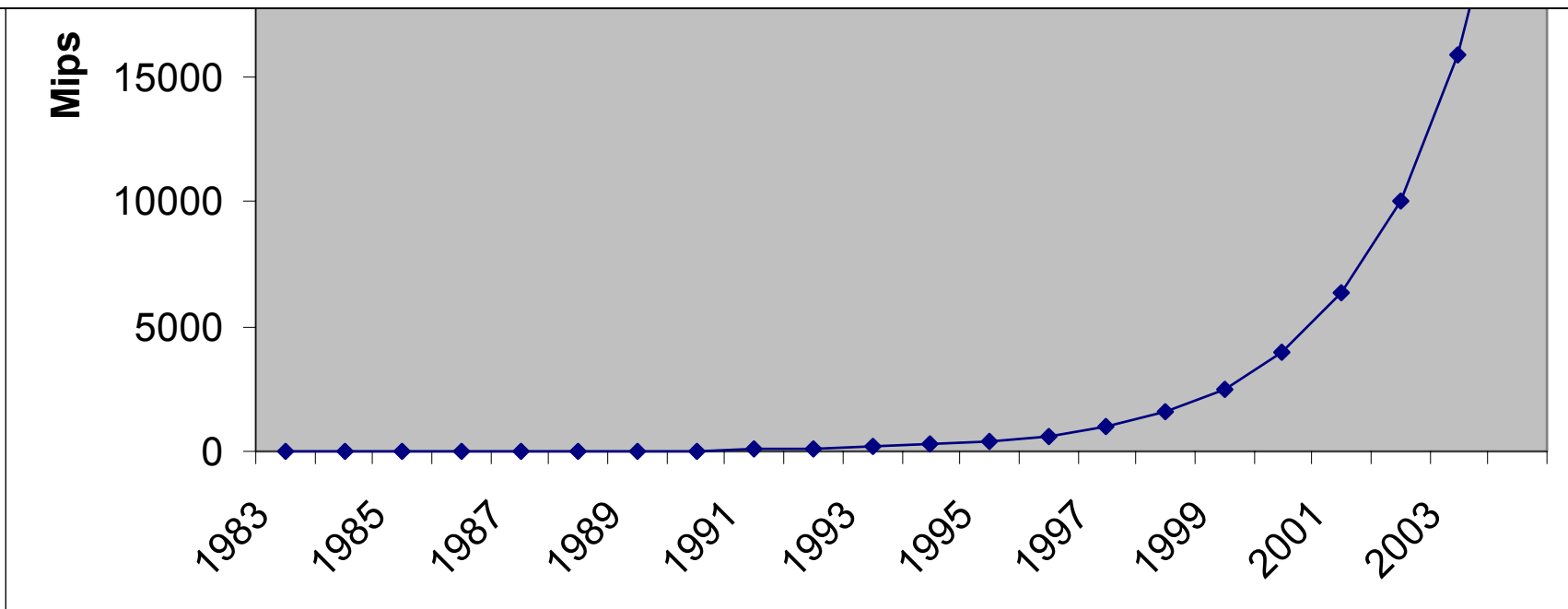
If you find 9 of them, you have a sense of observation above of the average.

If you find 10 of them, you are very observer.

If you find 11 of them, you are extremely observer.



- Als je in 1983 f 2500 p.m. zou verdienen: nu € 28,5 miljoen
- Als je auto toen 100 km/h zou rijden: door geluidsbarrière in 1989, door de snelheid van het licht in 2023
- Als het je toen een dag zou kosten om onregelmatigheden te vinden in het claimgedrag: nu 1 sec, oftewel 28.800 mogelijke cases per dag



Storage Price vs Time

Megabytes per kilo-dollar



Elke tien maanden verdubbelt de geheugencapaciteit.
1 Terabyte kost nu Euro 1000.



Profielen voor:

- Klantgedrag
- Aankoopgedrag
- Claimgedrag
- Particulieren
- Bedrijven
- Transactieverkeer
- i.h.a.: gecombineerd voorkomen van gebeurtenissen

Profielen voor fraudedetectie:

Volgens bronnen zou alleen al in Nederland de totale fraude bij verzekeringen 800 miljoen Euro kunnen bedragen.

Maar u weet dat waarschijnlijk beter.

Businesscase is eenvoudig en snel rond te krijgen. Gebruik rekenkracht en slimme methoden/mensen kan veel geld opleveren!

Bovenaan (No 1) van de lijst van ICT-doorbraakprojecten 2003:

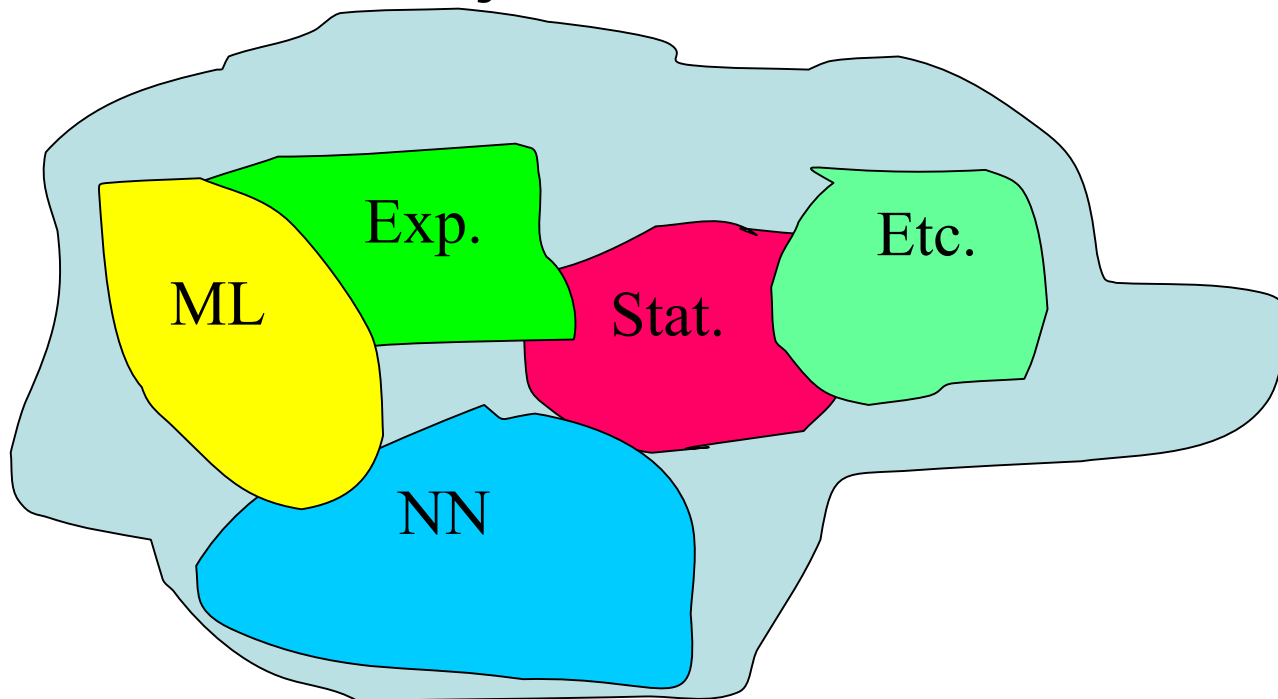
The main objective of the Diana-project is to develop new technology for building adaptive systems that intercept huge streams of data, analyze them in real-time and provide useful feedback.

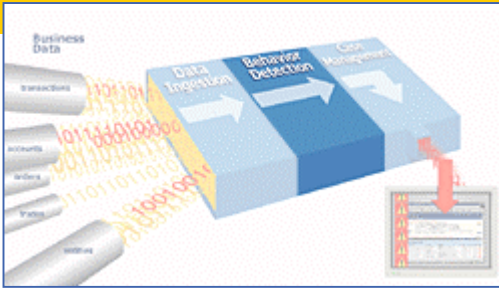
This technology can be used for building a new generation of systems for fraud detection with electronic payments, detection of misuse of complex information systems, network intrusion detection, gathering business intelligence, etc.

See Wojtek Kowalczyk, VUA

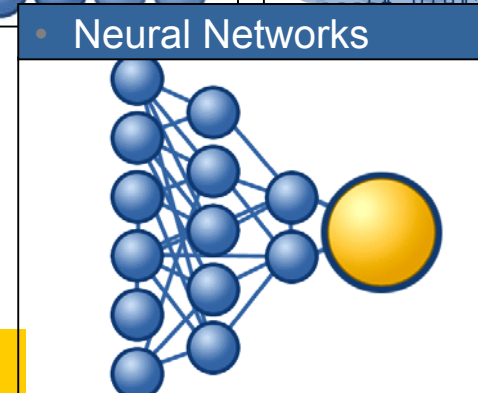
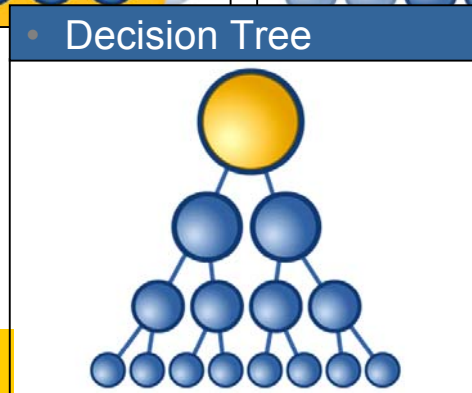
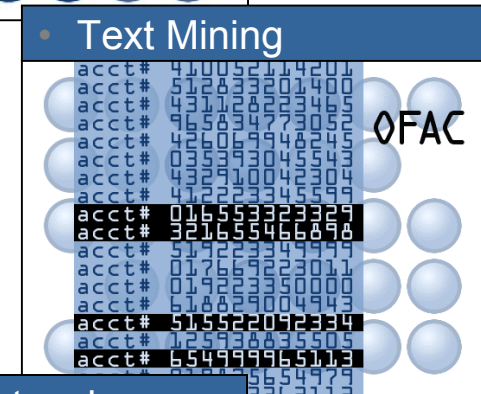
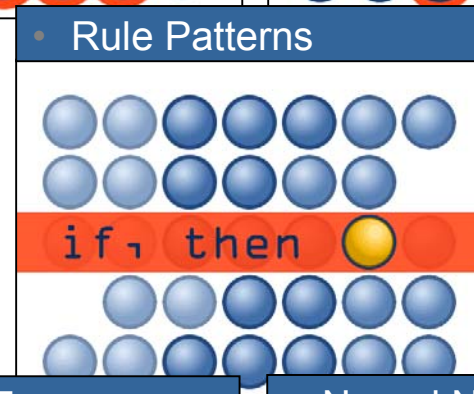
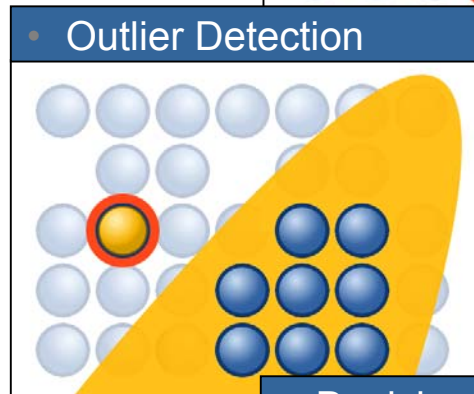
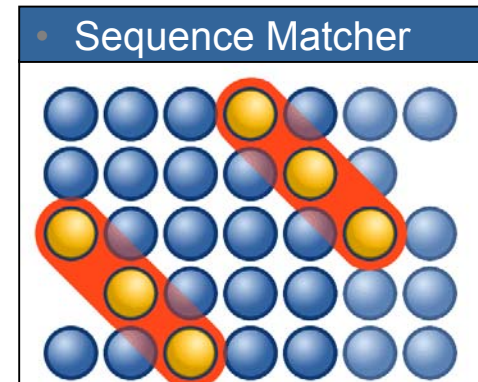
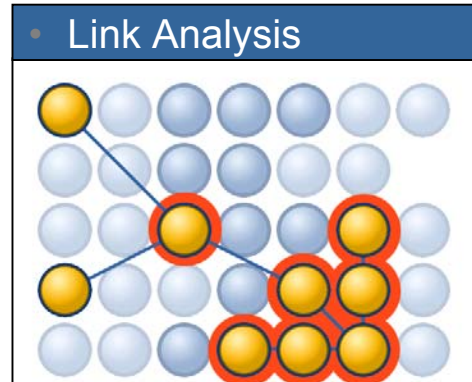
Meest succesvol zijn zgn. hybride modellen: mengsels van expertkennis, statistische regels, neurale netwerken en technieken uit machine learning.

Het gaat om het combineren van harde regels en modellen + ***het voortdurend alert zijn!***





Our platform taps into a combination of technologies to solve your business problems



Statistische modellen:

Regressie, Principale componenten analyse, Multi Dimensional Scaling, Multivariate discriminant analyse, Factoranalyse, CHAID, CART, ...

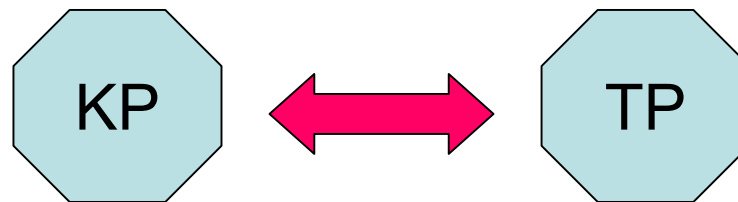
Technieken uit Machine Learning:

Rough Data Models, Naive Bayes, Decision Trees, Neurale Netwerken, Bayesian Netwerken, Support Vector Machines, Boosting, Bagging and Stacking,

Twee-componenten model

Een voorbeeld van een twee-componentenmodel vinden we bij betalingsverkeer:

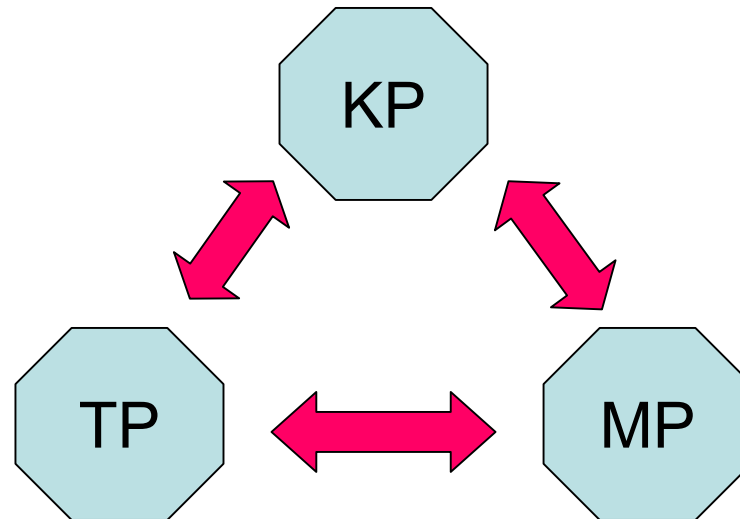
Klantprofiel en transactieprofiel



Drie-componenten model

Een voorbeeld van een drie-componentenmodel vinden we bij creditcards:

Klantprofiel, transactieprofiel en merchantprofiel

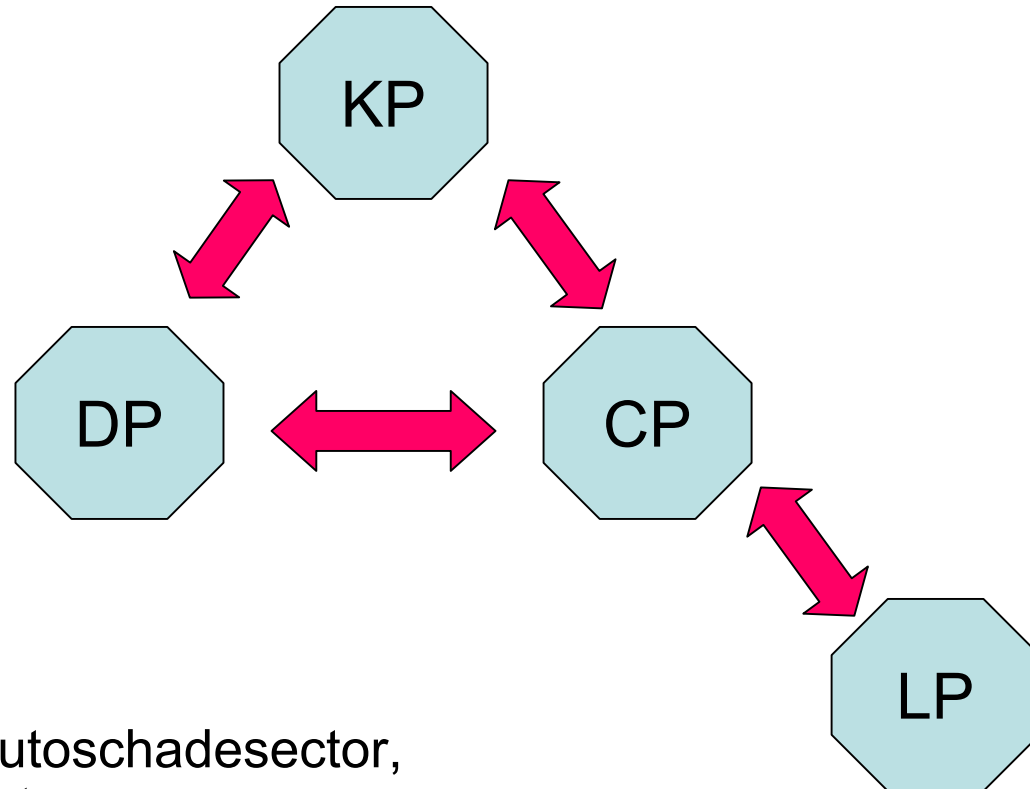


Idem in reisverzekeringen, opstalverzekeringen, inboedelverzekeringen, etc.

Vier-componenten model

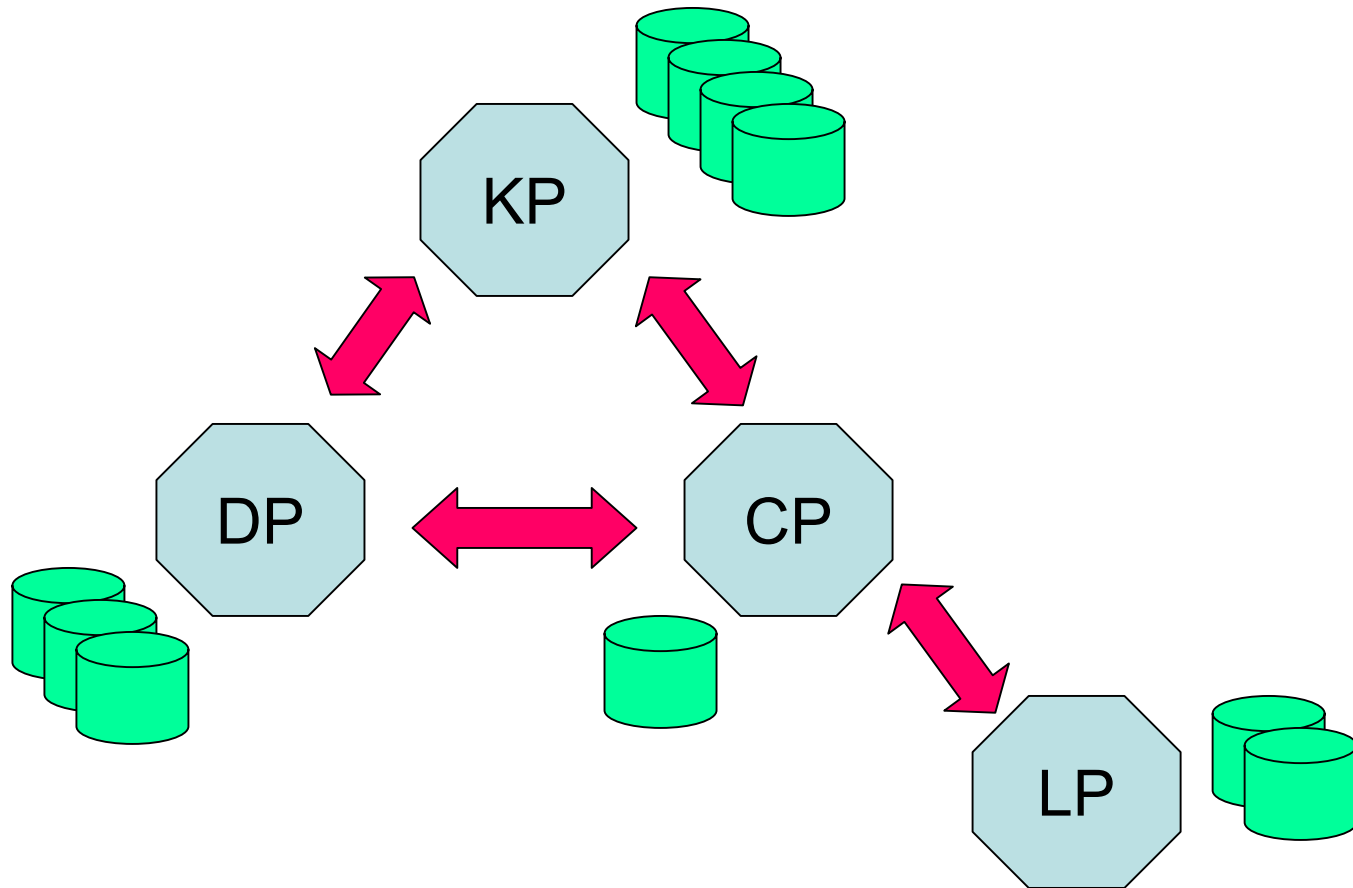
Een voorbeeld van een vier-componentenmodel vinden we in de gezondheidszorg:

Klantprofiel, claimprofiel, diagnoseprofiel en de leveranciersprofiel.

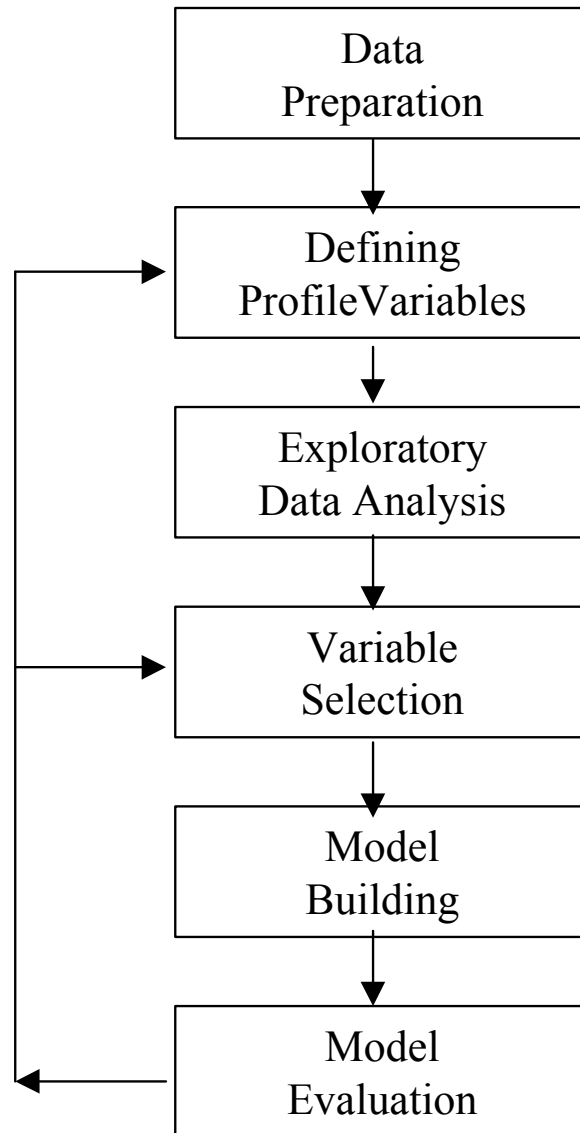


Maar ook in de autoschadesector,
huizenverkoop, etc.

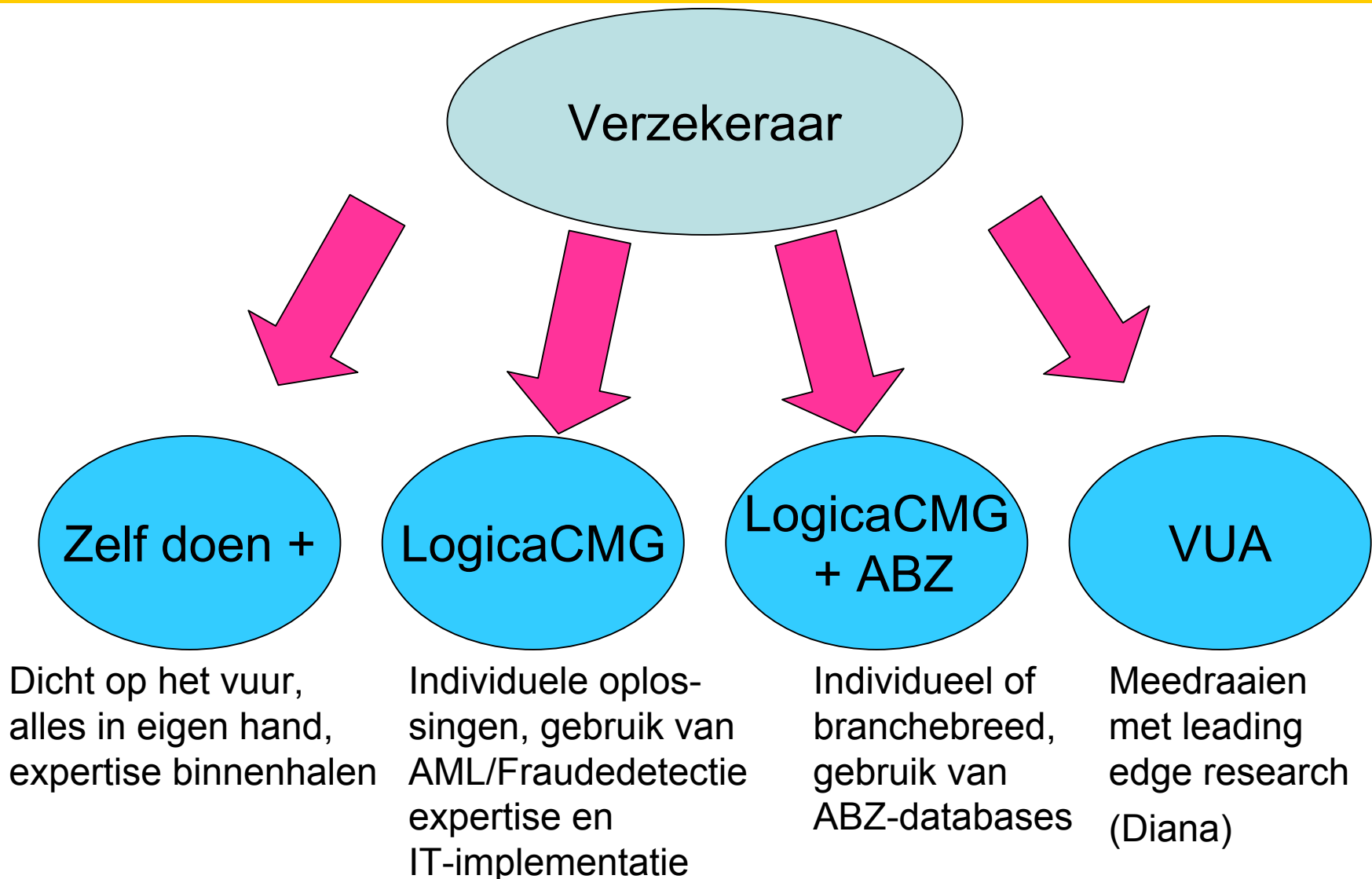
Profielen zijn slimme extracties



Veel data!! Complexiteitsreductie en zinvolle modellen



- Liefst zowel de “bad”-guys als de “good”-guys (0-en en 1-en).
- Er zijn nu ook succesvolle technieken die toepasbaar zijn op alleen de “bad”-guys.
- Het creëren van metafiles uit de enorme achterliggende bestanden is cruciaal.
- Aan cases wordt een score toegekend, bijv. tussen 0 – 100;
 - a. Hoe hoger de score, hoe “verdachter” de case
 - b. Sorteert van hoog naar laag
 - c. Bepaal drempelwaarde: kijk alleen naar cases met boven de waarde dan de drempel
 - d. Onderzoek en pas evt. de profileringsengine er op aan.



Vragen, opmerkingen en discussie